

Earth Syst. Sci. Data Discuss., author comment AC1  
<https://doi.org/10.5194/essd-2022-40-AC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Reply on RC1

Jamie Hannaford et al.

---

Author comment on "The enhanced future Flows and Groundwater dataset: development and evaluation of nationally consistent hydrological projections based on UKCP18" by Jamie Hannaford et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-40-AC1>, 2022

---

The title of the paper is misleading, it had me expecting to read about projected changes in flows and groundwater in the UK, when in fact the paper presents an evaluation of hydrological models prior to being forced with future runs. Indeed the opening sentence of the abstract doubles down on this - the paper presents a dataset of nationally consistent hydrological projections for the UK based on the latest UK Climate Projections. The paper doesn't do this and thus needs to be reframed as a paper evaluating hydrological models and RCM ensemble members in capturing observed regime - PRIOR to providing climate change projections for the sector.

Apart from this major oversight, the paper is well written and presented. There are a number of questions I was left with after reading that should be addressed in a revision. First, why the differential sampling of models for flows and groundwater? 3 different hydro models but only one GL model and one recharge model. In addition, why was uncertainty in model parameters not investigated.

>>> Thank you for the positive assessment of the paper's communication. The different choice of models reflect availability and suitability of models for both flow and groundwater. Parametric uncertainty was beyond scope, given the high number of existing model runs, and is highlighted as a priority in the future look (sect. 8)

Second, was any testing done on the transferability of models to future conditions? In particular the lumped conceptual models (GR and PDM) are held to perform better for modified catchments because the calibration picks this up. However, for future runs for later in the century are these modifications likely to be stationary and therefore how representative are these models for these catchments likely to be? On a related note, it would be useful to describe the range of disturbances affecting catchments - is it predominantly abstractions?

>>> This is always the perennial problem with future runs using calibrated models. However, the purpose of eFLaG is not really to look at such questions of non-stationarity in catchments, abstractions etc, which is a major topic in itself. Rather eFLaG is about running the models under current conditions and exploring future climate impacts. We suggest adding some caveats to the discussion to emphasise this point. Separately, a follow-up study is already looking at future changes in abstractions/discharges as well as

natural flows. We will highlight this avenue for future work in section 8. The range of disturbances are broad and users can explore these through the NRFA 'Factors Affecting Runoff' codes, and we will highlight this with a reference. Abstractions are prominent, but in some catchments discharges entirely balance or outweigh abstractions.

IF the purpose of the dataset is to inform adaptation planning why was RCP8.5 selected and why not more scenarios. There is debate in scientific literature presently about the realism of RCP8.5 - its selection here needs to be justified further. I understand the need for crystallising the uncertainty as stated, but this needs to be done based on the objectives or intended uses of the dataset.

>>> This is a fact of the climate projections – the regional projections are chosen as they are spatially coherent and transient but there is only one scenario. This is a limitation of that product. We do highlight this in the caveats, but will strengthen it and highlight the need to look at other projections. We will also make the fact clear in the earlier 'UKCP data' section.

Am I correct to conclude that there is no evaluation of model performances for an independent verification period, ie. that all metrics presented are for calibration period? How then can we be sure that calibrated models perform outside of the period used to train them, even during the baseline period, never mind under changed future climates. Please justify NOT using a verification period and what this may mean for your results.

>>>We could in theory have split the observed flows into a calibration and unseen validation period (used for calculating the statistics only) and then later recalibrated on the entire period to get the best calibration for use with RCMs. But, given we typically have many years in our calibration period, this would not make much difference. The paper by Harrigan et al (2018) using GR4J (see table 2 especially) shows this.  
hess-22-2023-2018.pdf

Minor comments

Some aspects of the uncertainties explored seem to be missing from Table 1 - eg. The first sentence of section 3 mentions 2 regional climate models, groundwater models are also not included.

>> this is just the wording, what is in the brackets is the GCM and RCM separately, we will reword to avoid ambiguity. We will add the groundwater to the table.

Line 196 can you say how this was done for clarity

>>not sure what is not clear, here we are saying what we are not doing (infilling to get to 365 like the Prudhomme 2012 study).

Line 216 - it would be useful to provide a brief overview of this downscaling method here. It is an important part of the methods and this is the go to paper describing the dataset.

>>>we omitted this for brevity since it is a standard method used in many previous studies and well described in these references. In response to this comment we will add the following: 'The bias-corrected precipitation products were then downscaled from 12km to 1km based on the distribution of the observed Standard-period Average Annual Rainfall (SAAR) over the period 1961-1990, as in previous studies (Bell et al., 2007; Kay & Crooks, 2014). This involved calculating the ratio of the observed SAAR at 1km to the observed SAAR averaged over 12km, and then multiplying RCM values by this ratio. This ensured that the spatial variability of rainfall was captured, but the total rainfall across the original 12km RCM grid cell remained unchanged.'

Line 236 - what does copied down mean here - lacks clarity.

>>Agreed, we will describe this more clearly, adding 'by simply setting all 1km grid cells to the value of the containing 12km grid cell.'

Figure 2 - there are 12 PPE members from text but 15 columns. Not sure I follow, is there an ensemble mean presented?

>>there are 12 columns, but the labelling is not 01 – 12 but 01 – 15, with some missing (e.g. 14). This is just the nomenclature of the RCM runs. We will make this clear in the caption.

I like how there is traceability between FFWGL and this dataset in terms of models, catchments etc. This is sensible and useful. I also like how research and industry needs feed into catchment selection. I also like the description of stage 2 assessment and its expectations. Don't often see this as nicely explained.

>>Thank you very much!

Line 604 - this concerns me and needs teasing out a bit more, especially given the objective of informing adaptation in the sector. What does 'to a degree' mean? I can understand the need to assess future changes relative to the 'real' present (ie disturbed) for planning but what are the implications and cautions that should be borne in mind?

>>>We agree that 'to a degree' is ambiguous and warrants fuller explanation. As highlighted above, that is exactly our objective, climate changes relative to the real (disturbed) present – as with the original, widely used FFWGL product, which also did not take account of human influences.

We will add a paragraph earlier in the methods, and then another sentence or two to this discussion, to highlight that this is an obvious simplification – entirely justifiable for most planning purposes, but one which needs to be considered in other applications where users may be interested in future changes in such influences. As noted, having G2G is an advantage as it does allow one naturalised run for comparisons.

Figure 4 caption needs to state #catchments and refer to table detailing the skill scores.

>>>agreed, will change

Line 644 missing word.

>>>agreed, will change

Line 680 full stop missing.

>>>agreed, will change

Line 840 - should really discuss the transferability of models and parameter sets from current to future conditions as a key source of uncertainty here, along with the stationarity of disturbances in relative catchments.

>>>agreed, as above, we will strengthen coverage of this point throughout the paper.